DOCUMENT RESUME

ED 333 755 FL 019 266

AUTHOR Kitao, S. Kathleen
TITLE Principles of Quantitative Research.
PUB DATE 91
NOTE 22p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Interpretation; *Educational Research; Foreign Countries; Language Research; *Research Design; *Research Methodology; *Research Problems

ABSTRACT

Research results should not be taken at face value; some research is not well designed, and readers must be able to assess whether the research carried out actually supports the results or may be explained otherwise. Research reports are usually divided into introduction or literature review, methods, results, and discussion and conclusions. Basic components of a study include variables and constructs, dependent and independent variables, and hypotheses. The reader must determine whether the components are clear and relationships are identified. It is also necessary to understand the concepts of variance and control and their application to research design. Potential threats to the external validity of a study include the length of treatment and sample selection. Threats to internal validity may include the following: participant selection or grouping; participant mortality (withdrawal); effects that may occur simply in the passage of time; participant and researcher attitudes; validity (face, content, and construct) and reliability (stability and equivalence) of the measures used, and the effect of practice on results; and designs that lead to different treatment of groups or do not eliminate alternative explanations. It is also important to realize that not all threats to validity can be avoided, but can be acknowledged and controlled. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED333755

Principles of Quantitative Research

S. Kathleen Kitao

Doshisha Women's College

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Kitao

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

FL019206

Language teachers, like members of many other professions, need to read research papers in order to keep up with developments in their profession. Research is, in fact, expanding rapidly in the field of language teaching, so the ability to read and understand research is becoming increasingly important (Dean, 1988). However, the results of studies should not necessarily be taken at face value. Some of the research that gets published is not well designed, and these problems in research design may allow for explanations of the results that are not considered by the researchers. Therefore, readers of research need to have the ability to evaluate the research design to see whether the research done actually supports the results or whether there may be other explanations for the results.

Research papers are usually divided into four sections--the introduction or literature review, the methods section, the results, and the discussion and conclusions. It is tempting, in reading research papers, to read the first and last sections most closely. They are the most interesting sections, the ones most likely to have useful information. The methods and results sections do not seem very important, so in the interest of time, they may be skipped altogether. However, this kind of attitude is actually dangerous--without reading and carefully evaluating the methods section, readers do not know whether the research design has problems that allows for rival explanations of the results. Readers cannot judge whether what the researchers claim they found is

actually supported by the results of the study (Dean, 1988). In this paper, I am going to discuss concepts related to research design and then potential problems of research design and how they might be overcome.

It should be noted that this paper deals with quantitative research, that is, research that depends on expressing constructs in numbers and makes use of statistical analysis of those numbers. (Qualitative research, in contrast, depends more on the subjective observations of the researchers.) Qualitative research, while interrelated with quantitative research, is based on different principles, so the principles discussed in this paper are not intended to be applied to the evaluation of qualitative research.

Some Definitions

In order to evaluate research design, it is necessary to clearly understand the basic components of a research study. In this section, I will discuss definitions of various components and distinctions made among them.

Variables and Constructs

The most basic component of any study is its variables. A variable is, most simply, something that varies. It is a value that changes over time or differs among individuals. Some examples of variables are gender, income, and IQ scores.

Dependent and independent variables. A crucial distinction made among variables is between independent and dependent variables. Independent variables are variables that are hypothesized to be a cause or to have an influence on another variable. Dependent variables are

variables that are hypothesized to be an effect or to be affected by another variable.

Variables are not inherently dependent or independent; how a variable is classified depends on how it is linked to other variables in a hypothesis. (A hypothesis is the statement of relationships among variables.) For example, if the hypothesis is that students taught using method X will achieve higher scores on a language proficiency test than students taught using method Y, the method is the independent variable and the language proficiency score is the dependent variable. Conversely, if the hypothesis is that students with high language proficiency scores will attain a higher grade point average in an American university than students with low scores, then the language proficiency score is the independent variable and the GPA is the dependent variable. Therefore, a particular variable can be either independent or dependent, depending on the hypothesis.

Control variables. Another type of variable is control variables. Control variables are extraneous variables that must be kept constant, or controlled, because they may have an effect on the dependent variable. For example, if the hypothesis states that students with higher language proficiency scores will attain a higher GPA, it might be necessary to control for IQ scores, since IQ might also influence GPA.

Constructs and variables. In considering research, it is important to understand the relationship between a construct and a variable. A variable is essentially what we can observe or quantify of characteristics or

abilities, while a construct is the underlying characteristic itself. For example, English proficiency is not something that can be seen or measured directly. However, it can be measured indirectly using various instruments, and the results of these instruments are variables. The variable for the construct of English proficiency might, for example, be a TOEFL score or some other language test score. It is important to keep in mind, however, that though the TOEFL score is intended to reflect language proficiency, it is not language proficiency. In evaluating research, it is necessary to consider whether the measures used for the variables do actually reflect the underlying construct they are intended to reflect.

Therefore, an important issue related to the difference between a construct and a variable is operationalization. Operationalization is the link between the construct and the way it is measured; the operational definition of a variable is the way that it is measured or quantified. For example, English proficiency might be operationalized as a TOEFL score, a rating of a person's ability to carry on a conversation in English under certain conditions, or the results of a reading test.

There are two requirements for an operational definition. It must be conceptually linked to the construct that it is intended to represent, and it must be unique to that construct, that is, it must measure that construct and no other. For example, a test of listening proficiency that requires some reading may also be

measuring reading proficiency, that is, the score for listening proficiency may be influenced by the participant's ability to read. In that case, the operational definition includes a construct that the researcher did not intend to measure.

Operational definitions are important for three reasons (Hatch and Lazaraton, 1991). A clear operational definition allows the researcher to be consistent throughout the research; it allows readers of the research to clearly understand the findings; and it allows other researchers to replicate the research.

In considering the relationship between variables and constructs, it is necessary to consider the definitions that the researcher uses, if they are specified. Whether a variable truly reflects a construct is, in part, dependent on how that construct is defined. For example, if the definition of English proficiency is related to the ability to communicate in English, then a TOEFL score is not a good operationalization.

Hypotheses. In quantitative research, unless the study is a pilot study, the expected relationship among the variables will be specified in the form of one or more hypotheses, which are tested in the study. The researchers should give logical arguments or cite previous studies to support these hypotheses (Personal communication, James M. Stiff, 1986).

Questions about the basics of a study. In order to understand a study clearly, it is helpful to ask several questions about its variables and constructs, and their operationalization. What are the constructs and variables

involved in the study? Are they clearly identified and defined? Are the operational definitions of the variables clearly linked to the constructs they are intended to represent and not to any other constructs? Is the relationship that the researcher has specified among the variables supported by previous literature and/or logical arguments?

Variance and Control

Two more concepts that are basic to an understanding of research are variance and control. In this section, I will discuss these concepts and their application to research design.

The goal of research design is testing hypotheses. In order to test hypotheses, two things are necessary--variance and control. Variance is having different values for the dependent variables. For example, if all participants in a study are approximately equal in language proficiency, it is impossible to test a hypothesis related to language proficiency. Control is the ability to explain as much of the variance as possible. The more variance there is that cannot be explained, the less control there has been in the study. Lack of control allows for rival explanations for the results.

Types of Variance

Variance can be divided into three types--systematic experimental variance, extraneous variance, and error variance. Systematic experimental variance is differences in dependent variables caused by the experimental manipulation and statistical control (that is, the

effect of the independent variable). This is the type of variance that researchers want to maximize. Extraneous variance is caused by unspecified and unidentified variables. Error variance is variance caused by sloppy experimental procedures, errors in measurement, coding errors, key punching errors, etc. Obviously, the latter two types of variance should be minimized.

In quantitative research, the purpose of research design is to achieve control--to minimize the extraneous variance. Researchers must try, as much as possible, to either keep variables (other than the independent variables) constant that might influence the dependent variables or measure them and control their influence statistically.

Threats to Validity

Internal and External Validity

With this background, I will discuss some of the threats to the validity of the results of research. In general, these threats can be divided into two categories--threats to internal validity and threats to external validity. Internal validity is the degree to which we can attribute variance to the manipulation. This concept is related to control, discussed in the last section, in that threats to internal validity are factors that reduce control and increase extraneous variance. External validity is the extent to which the results of the study can be generalized.

Threats to External Validity

In considering the external validity, the reader of a study must consider whether the conditions found in the

study have any relationship in the "real world." For example, a study of the effectiveness of a language teaching method may involve using the method only once or only over a short period of time. Since, for the most part, language classes use a method over a much longer period of time--possibly several months--there have to be questions about the external validity of such a study.

The selection of participants in the study may also have an influence on the extent to which the results of a study can be generalized. If the participants in a study are selected from non-native speakers studying English in the United States, it is difficult to generalize the findings to students of English who live in countries where English is a foreign language.

Threats to Internal Validity

Threats to internal validity can be divided into several categories according to the issues that they are related to. In this section, I will discuss each of these categories and the problems related to them.

Grouping Issues. The first category is grouping issues. These are threats to validity related to the ways that participants for a study are chosen and grouped, as well as their continued participation, and how these influence the results of the study.

In quanitative research, it is ideal to select participants at random. This helps assure that the sample reflects the population that it is intended to represent. (The sample is the actual participants in the study; the population is the group of people that the study is intended to be generalized to. For example, the sample

may be a group of randomly chosen students at an American college; the population may be all American college students.) However, it is not always possible to randomly select participants, and so researchers often resort to other ways of choosing participants, including self-selection and using intact groups. Each of these represents a threat to internal validity.

Self-selection is the common practice of having participants volunteer to join the study. This is particularly a problem when participants who volunteered are compared to participants who did not (e.g., who participate in the study as part of a class). Obviously, participants who volunteer to participate in a study may be different from people who don't. For example, in a study of a language learning method, if volunteers are used to try the method in question, the volunteers are likely to be more motivated and may have a higher language aptitude and a higher proficiency than non-volunteers. Therefore, researchers must avoid comparing volunteer and non-volunteer groups, and they have to be cautious about generalizing from volunteers to the general population. While it is desirable to avoid self-selection, it is not always practical--researchers are not always in a position to require randomly chosen participants to participate in a study. Therefore, if self-selection is used, researchers need to acknowledge the self-selection and its possible effects on the results, taking particular care in comparing volunteer and non-volunteer groups and generalizing from a volunteer sample to the general population.

The use of intact groups is also a potential problem. This is the practice of using existing classes or other existing groups as participants in a study. There are two major problems with this practice. One problem is that the groups may differ in some variable, since participants have not been randomly assigned to the groups. One group might, on average, have a higher language proficiency or might have spent more time in English-speaking countries. If so, the groups are not necessarily comparable and differences between them on the dependent variables may be explained by preexisting differences in the two groups rather than the independent variables. In addition, the different intact groups may have already established ways of interacting among themselves, and this may influence the dependent variable as well. This would be a particular problem in doing research on, for example, a teaching method, because the interaction among the members of the class may influence the effectiveness of the method.

Researchers should avoid intact groups if possible and instead assign participants to groups randomly. However, in many cases, this is not feasible, and in such cases, there are ways to minimize the resulting problems. Researchers can identify other variables that might influence the dependent variable, measure them, and use statistical procedures to determine whether they have influenced the outcome and to control them.

Another issue related to the participants is mortality. Mortality is the dropping out of participants from the study for any reason. This is a problem because

the participants who drop out may be different on some variable than those who do not. In addition, mortality can change the mean of the group on some characteristic. Researchers should specify how many participants dropped from the study and why. If possible, they should try to find out whether these participants were different on any important variables from participants who did not drop out.

Time-Related Effects. Another category of threats to validity is time-related effects. These are related to the passage of time between the beginning and end of a study.

One type of time-related threat to validity is history effects. History effects are the effects of incidents not related to the study that intervene between the beginning and end of the study and which may have an effect on the results of the study. For example, a current event might intervene in a study of changes in attitudes over time and may effect the attitudes in ways that are not related to the researcher's manipulations or any variables that are being considered. In order to be aware of history effects, researchers should try to have sources in the setting where they are researching in order to find out about any possible history effects, and, if possible, they should talk to research participants as well. Having control groups also helps, since the history effect would presumably have an influence both groups, and any difference between the two could be attributed to the treatment. (However, control groups do not eliminate the problem entirely, because it is still possible to have an

interaction between the treatment and the history effect.) Finally, the study should not be continued any longer than necessary. (There may, however, be other reasons continue a study over a long period of time, so researchers have to find a balance between the advantages of having a long study and those of keeping the study short.) This is obviously a very important issue in the area of language teaching methodology, since external validity depends partly on using the method over time.

Another type of time-related effect is maturation effects. Maturation effects are normal changes in participants over time. In studies related to language teaching, these changes are probably most obvious in children, since their muscle control, their competence in their own language, etc., are still maturing, but maturation can be a factor in studies with adults, too. In a study done with non-native English speakers living in an English-speaking country, the fact that participants are in an English-speaking environment will certainly influence their language proficiency, independent of the manipulation in the study. Again, having a control group and keeping the study short help limit the effects of maturation or at least allow researchers to recognize them. In addition, researchers may use two pretests, a treatment, and then a posttest, and then compare the two pretests in order to find out if there is maturation occurring.

People-Related Effects. A third category of threat to validity is people-related issues. This category is related to the attitudes of those involved in the study,

both participants and researchers.

One people-related effect is called the Hawthorne Effect. In some cases, just the fact of being involved in a study and/or the novelty of the treatment itself can cause participants to perform better than they normally would. Therefore, when researching the effectiveness of a new language teaching method, researchers have to take into account that the very novelty of the new method might have an effect on the results, not the characteristics of the method itself.

Another type of people-related effect is participant expectancy. If participants understand what they study is about or what the researchers expect of them, they may consciously or unconsciously try to give the researcher the effect that he/she is looking for. To the extent possible, subjects should be blind to the purpose of the study and the hypotheses being tested.

Another type of people-related effect is researcher/coder expectancy. If the people administering the treatment or rating the results are aware of the purpose or hypothesis of the study, this may have an effect on the results. For example, if participants are divided into high and low aptitude groups, and if the people administering the treatment are aware of which groups have high and low aptitudes, it may subtly influence how they treat the groups. Similarly, if they are aware of the hypotheses, they may, without intending to, influence the responses of the participants. In studies comparing different methods, the researchers themselves often teach the classes that are intended to show that one method is

more effective than another, and they may influence the results. If coders are aware of the hypotheses, there may be an influence on the way that they code the results of the study. As with participant expectancy, these problems can be dealt with by having the people who code the results or administer the treatments blind to the hypotheses and to the group assignments.

Measurement Issues. Another category is measurement issues. This includes issues related to the instruments used to measure the constructs in the study (tests, questionnaires, rating systems, interview protocols, etc.) and the ways they are administered. In the area of measurement issues, there are two components where there may be threats to the validity of the results of the research. The validity and reliability of the measures themselves may effect the validity of the conclusions drawn. (I will refer to the validity of the measurement instruments as measurement validity to distinguish it from the validity of the results of the study.) In addition, procedures for administering the measures may effect the results. These can all affect the validity of a study.

Measurement validity is the extent to which the instrument measures what it purports to measure. There are three types of measurement validity. Face validity is the extent to which a measure "looks" like it measures what it is supposed to measure. This is the logical connection between the construct and the instrument that is intended to measure it.

The second type of measurement validity is content validity, which is the extent to which an instrument

measures all the facets of the construct. This is a conceptual problem, not a statistical problem, and it depends to a great extent on how the construct is defined. If a measure that is purported to measure communicative ability in a second language only tests listening ability, it is not measure all facets of communicative ability.

The third type of measurement validity is construct validity. It is related to correlations between the variable in question and other variables. There are three types of construct validity. The first is convergent validity, which means that the variable is correlated with variables that it should be correlated with. The second is discriminant validity, which means that the variable is not correlated with constructs is should not be correlated with. For example, measures of lexical knowledge and reading proficiency might be expected to be correlated (convergent validity), but measures of reading proficiency and communication anxiety would not be correlated (discriminant validity).

Reliability, the second aspect of measurement, has two aspects. One is stability, that is, consistency over time. For a construct that has not changed, there should be a high correlation between a participant's scores at two points in time. One problem related to the assessing the stability of a measure is separating true change from lack of reliability. Stability is also evaluated by comparing the results of alternate forms of a measure.

A second aspect of reliability is equivalence, which is the extent to which all items reflect the same construct. For example, if a researcher is measuring

attitude toward the target language, responses to all of the questions should reflect a similar attitude. If the answer to one question seems to reflect a different attitude, it may be measuring some other construct. Cronbach's alpha is a statistic used to measure equivalence, and if the alpha values reported for the questions in a measure are low, the researcher should consider removing or revising one or more questions.

The practice effect is a third measurement issue. If more than one measurement instrument is used in a study, or if the same instrument is used more than once, the results of the instruments administered later may be affected. For example, if the same test is administered at the beginning and end of a study, having taken the test once already may help the participants to do better on it a second time. While the use of equivalent rather than identical measures helps to some degree, it does not eliminate the problem entirely, since taking the first measure may make participants aware of issues, grammar points, etc. Again, a control group which responds to the measures but does not receive the manipulation will help eliminate this problem. In addition, if more than one measure is used, it is wise to counter balance the measures (administer one measure to half the participants first and have the other half take the other measure first). Statistical methods can be used to show whether there was a practice effect.

Design Issues. Another category of threats to validity is design issues. This category includes research designs that lead to unintentionally different

treatment of the groups and research designs that are inherently faulty because they do not eliminate alternative explanations.

The first category of design issues is uncontrolled differences in the treatment of the groups. For example, in a study of different language teaching methods, if different teachers are teaching classes using different methods, the possibility must be considered that differences found in improvement in participants' language proficiency may be due to differences in the teacher, not differences in the method. This problem should be avoided, if possible, by having all classes taught by one teacher or by having each teacher involved teach classes using two or more methods. Another type of problem in this category would be in a study in which participants meet outside of regular classes for instruction using a different method. If their improvement is then compared to the improvement of students in regular classes, there are must be questions about the validity of the results. In addition to self-selection problems (comparing volunteers and non-volunteers), participants who meet outside of regular class time are receiving more hours of instruction than those who are only in regular classes. This difference may account for any differences in their improvement.

The second category in design issues is faulty research designs. The three designs discussed below are inherently flawed and do not allow any conclusions to be drawn about the sources of the variance (Hatch and Lazaraton, 1991).

One faulty design is called a one-group posttest only design. In this design, only one group is used, and there is a manipulation (or series of manipulations) followed by a posttest. Based on the results of such a study, no claims can be made about the effect of the manipulation. Since nothing is known about the levels of the variables before the manipulation, it cannot even be claimed that they changed, much less that they changed in a particular direction or as a result of the manipulation.

A second faulty design is the one-group pretest/ posttest design. In this design, again, only one group is used, and there is a pretest, a manipulation, and a posttest. In this case, the results of the pretest and the posttest can be compared, so it can be determined whether there were changes in the values of the variables. However, it is not possible to draw conclusions about the source of the changes. It may have been the treatment. It is also possible that there were history effects, maturation effects, or practice effects that explain any variance that was found.

The third type of faulty design is a two-group, no control design. In this type of study, there is an experimental group (which experiences the manipulation) and a nonequivalent control group, with a pretest and a posttest. Therefore, it is impossible to know whether the response to the treatment is due to particular characteristics of the experimental group and whether the results can be generalized beyond that group.

Conclusion

In assessing threats to validity, it is necessary to recognize that not all threats can be avoided. On one hand, there are practical limitations. Researchers may, from a practical point of view, need to choose between using volunteer groups and using intact groups. There may be no other alternative.

In addition, in designing a study to avoid one type of threat, the researcher may be unavoidably opening the study to another type. Finding a balance between internal validity and external validity is a particular problem. Avoiding threats to internal validity is likely to increase threats to external validity. Situations that provide the control necessary to avoid extraneous variance may not generalize well to other situations.

Also, there are advantages both to keeping a study short (for example, avoiding history and maturation effects) as well as to making it longer (avoiding the Hawthorne Effect and practice effects). Again, researcher need to try to find the best balance among these factors.

In all these cases, researchers should acknowledge threats to validity, control for them as much as possible, and discuss the influence they may have had on the outcome of the study.

This has just been a very brief overview of the principles of quantitative research design. However, it is important for anyone who reads quantitative research to understand these principles in order to be able to evaluate what they read and make judgments about the validity of the results of the research.

List of References

Dean, J.D. (1988). Understanding research in second language learning. Cambridge: Cambridge University Press.

Hatch, E., & Lazaraton, A. (1991). The research manual: Design and statistics for applied linguistics. New York: Newbury House.